Building Molecular Charge Distributions from Fragments: Application to HIV-1 Protease Inhibitors

L. YOUNG,[1]* I. A. TOPOL,[1] A. A. RASHIN,[2] and S. K. BURT[1]

[1]Structural Biochemistry Program, Frederick Biomedical Supercomputing Center, SAIC Frederick, NCI-FCRDC, Frederick, MD 21702; and [2]BioChemComp Inc., Teaneck, NJ

Received 6 February 1996; accepted 25 June 1996

ABSTRACT

Interaction energies are a function of the molecular charge distribution. In previous work, we found that the set of atomic partial charges giving the best agreement with experimental vacuum dipole moments were from density functional theory calculations using an extended basis set. Extension of such computations to larger molecules requires an atomic partial charge calculation beyond present computational resources. A solution to this problem is the calculation of atomic partial charges for segments of the molecule and reassociation of such fragments to yield partial charges for the entire molecule. Various partitions and reassociation methods for five molecules relevant to HIV-1 protease inhibitors are examined. A useful method of reassociation is introduced in which atomic partial charges for a large molecule are computed by fitting to the combined electrostatic potential calculated from the fragment partial charges. As expected, the best sites for partitions are shown to be carbon—carbon rather than carbon—nitrogen bonds.

Introduction

A major challenge in the field of drug design is the calculation of free energies of binding. Since most drugs and inhibitors are not covalently bound to their targets, a good model of the electrostatic interaction is important. Atomic partial charges may be used to model this interaction; however, calculation of these charges must be performed such that they represent the molecular charge distribution as accurately as possible.

Semiempirical methods may be used to calculate these charges; however, in the case of certain small molecules, the molecular dipole moments from such computations do not agree well with experiment.[1] Calculations of molecular dipole moments using density functional theory (DFT) and an extended basis set with diffuse functions (DZVPD[2]) have shown exceptional agreement with

*Author to whom all correspondence should be addressed.

CCC 0192-8651 / 97 / 040522-11

experiment.[3,4] In addition, calculation of hydration enthalpies for 50 small molecules using partial charges fit to the DFT/DZVPD electrostatic potential agreed with experimental enthalpies of hydration to within 1.5 kcal/mol, on average.[4]

Extending the quantum-mechanical partial charge calculation to larger molecules such as HIV-1 protease (HIV PR) inhibitors (as large as 60 nonhydrogen atoms), while maintaining the large basis set, is beyond present computational resources. The usual solution to this problem divides the molecule into fragments of a size suitable for calculation and then reassociates these fragments, yielding partial charges for the large molecule. In the following sections, problems surrounding the fragmentation and reassociation of molecules for partial charge calculations are investigated using compounds relevant to HIV PR inhibitors. We introduce a novel method for reassociation based on the electrostatic potential. Finally, such partial charges for an inhibitor of HIV PR are used in a calculation of an inhibitor–enzyme interaction energy.

Fragment Problem

The use of a large basis set to calculate atomic partial charges for large molecules (50–60 nonhydrogen atoms) is at present computationally intractable. However, such a calculation may be performed for fragments and the fragments reassociated to yield the large molecule atomic charges. If atomic charges could be calculated for the molecule as a whole, the values would reflect intramolecular polarization effects. When the partial charges are calculated from combining fragments, some of this polarization effect is lost. Thus, the proper modeling of polarization is the issue. This issue may be addressed both by examination of different bond type partitioning and by the study of different fragment reassociation methods.

When partitioning the molecule into fragments, the best schemes may be those that cut bonds with the least polar character such as those formed between tetrahedral carbon atoms. For molecules in which such bond types are lacking, other fragmentation sites must be examined. Once a fragment is produced, a terminal chemical group must be added to saturate the cut bond (Fig. 1). To curtail large alterations of the fragment dipole moment, a near neutral terminal group, such as a hydrogen atom or methyl group, is used.

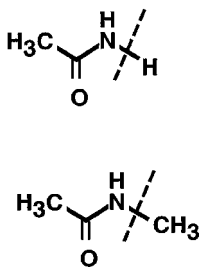

FIGURE 1. —H and —CH_3 terminal groups are shown saturating the cut bond (dashed line) of one of the fragments generated from a larger molecule. Methods of reassociation of the fragments must eliminate these terminal groups.

For reassembly of the fragments, two methods are tested here. (1) The neutralization method[5] simply sums the partial charges of the terminal groups into the atoms to which they are bonded. For example, the partial charge of the terminal H atom or terminal methyl group is summed into that of the adjoining nitrogen in Figure 1. (2) The CEP (combined electrostatic potential) fit method combines the fragment electrostatic potentials calculated from the fragment partial charges, including those of terminal hydrogen atoms. These values are calculated at the surface points of the whole molecule. The atom-centered partial charges for the reassembled molecule are calculated by fitting to this combined electrostatic potential. In the process, the partial charges of the terminal hydrogen atoms are eliminated and their effect is represented in the values of the new set of atom-centered charges.

In summary, the CEP fit method attempts to model the polar character of the fragmented bond using the extra hydrogens as part of the charge distribution located along the bond, in much the same way that bond-centered charges are sometimes used to better reproduce electric multipole data. The charges fit to the combined electrostatic potential are, however, only the atom-centered charges. In this way, the CEP fit eliminates the partial charges on the bond, representing their effect in the atomic-centered partial charges of the whole molecule. Although fitting the electrostatic potential to derive atomic partial charges is common,[1,6,7] using combined fragment electrostatic potentials to eliminate terminal hydrogens and thus yield charges for the reassociated molecule has not, to our knowledge, been reported.

A second version of the CEP fit method was also tested. Instead of combining the fragment

electrostatic potentials by calculating them from the fragment partial charges, the fragment DFT electrostatic potentials may be calculated at the molecular surface of the whole molecule, allowing a simple sum of the DFT fragment electrostatic potential values at each whole molecule surface point to produce the CEP for the reassociated molecule. This version has the advantage of eliminating the need for fragment partial charges.

Finally, two other methods are discussed. Overlap partitioning involves a different method of fragmentation. The fragments are cut such that they have several atoms in common. Upon reassociation these atoms overlap. Thus, in the average overlap method of reassociation, the overlapping atom partial charges are simply averaged.[8]

The other method is the use of neutral blocking groups.[9] This method allows any type and size of terminal group. Upon calculating fragment charges, this group is constrained to be neutral, allowing reassociation of the fragments by simply dropping the terminal groups.

Calculations

Fragment atomic partial charges were calculated using the electrostatic potential fit option[1,6] in Gaussian 92/DFT single point calculations.[10] This software was also used to calculate the fragment DFT electrostatic potential on the molecular surface of the whole molecule. The density functional calculations were performed using Becke's three-parameter exchange functional[11] with the Perdew correlation functional[12] using a DZVPD basis set.[2,13] To compare the recombined molecule charge distribution with the whole molecule charge distribution, values of the electrostatic potential (EP) from both were calculated at surface points.[1] Root mean square (RMS) differences in the EP values were used to evaluate the methods of reassociation of the fragments. This evaluation was chosen over a direct comparison of atomic charges, because the value of the interaction energy depends on the EP, and the set of atomic partial charges that can produce a given EP is not unique. Although the best results were expected for bonds cut between tetrahedral carbons, other types of bond partitions were also evaluated by EP RMS differences. In addition to the EP evaluation, additional properties were calculated, namely, dipole moments, free energies of hydration, and interaction energies.

Dipole moments were calculated from the atomic partial charges. Free energies of hydration of the test systems were calculated using the boundary element method.[4] The evaluation of the interaction energies of the enzyme–inhibitor complexes were performed using CHARMm 22[14,15] and AMBER 4.0,[16] with the corresponding all-atom force fields. Hydrogen atoms were added to the enzyme and minimized in energy to a gradient tolerance of 0.1 kcal/mol · Å. The nonhydrogen atoms were fixed.

Results and Discussion

Evaluation of various partitions and reassociations requires a comparison of the whole and reassociated molecules. For this reason, the molecules chosen must be large enough for fragmentation but small enough for a DFT/DZVPD electrostatic potential calculation. In some cases, more than one combination of partitions is tested (hereafter, a combination of partitions for a molecule will be referred to as the fragmentation scheme). The types of bonds partitioned or cut are mostly different types of carbon—carbon and carbon—nitrogen bonds. The types of atoms at the partitions are tetrahedral carbons (CT), carbonyl carbons (C), aromatic ring carbons (CR), proline alpha carbons (CP), and amide nitrogens (N). Thus, a fragmentation scheme with two partitions, each containing tetrahedral carbons, would be represented by the notation CT/ CT, CT/ CT, where the slash represents the cut bond.

We chose to study five model systems (A-E) as shown in Figure 2. The partitioning schemes studied are represented by dashed lines in Figure 2. These molecules are portions of HIV PR inhibitors and their geometries were taken from the enzyme–inhibitor crystallographic data.[17–19]

EP RMS COMPARISONS

The results from the EP RMS difference calculation between whole and reassociated molecules are given in Table I. In general, the smallest EP RMS differences are for the CEP fit method with partitions between carbon atoms. The italicized EP RMS difference values in Table I denote the smallest differences for each molecule. These range from 0.9 to 1.4 kcal/mol.[20] More specifically, the smallest difference in EP RMS for each molecule using

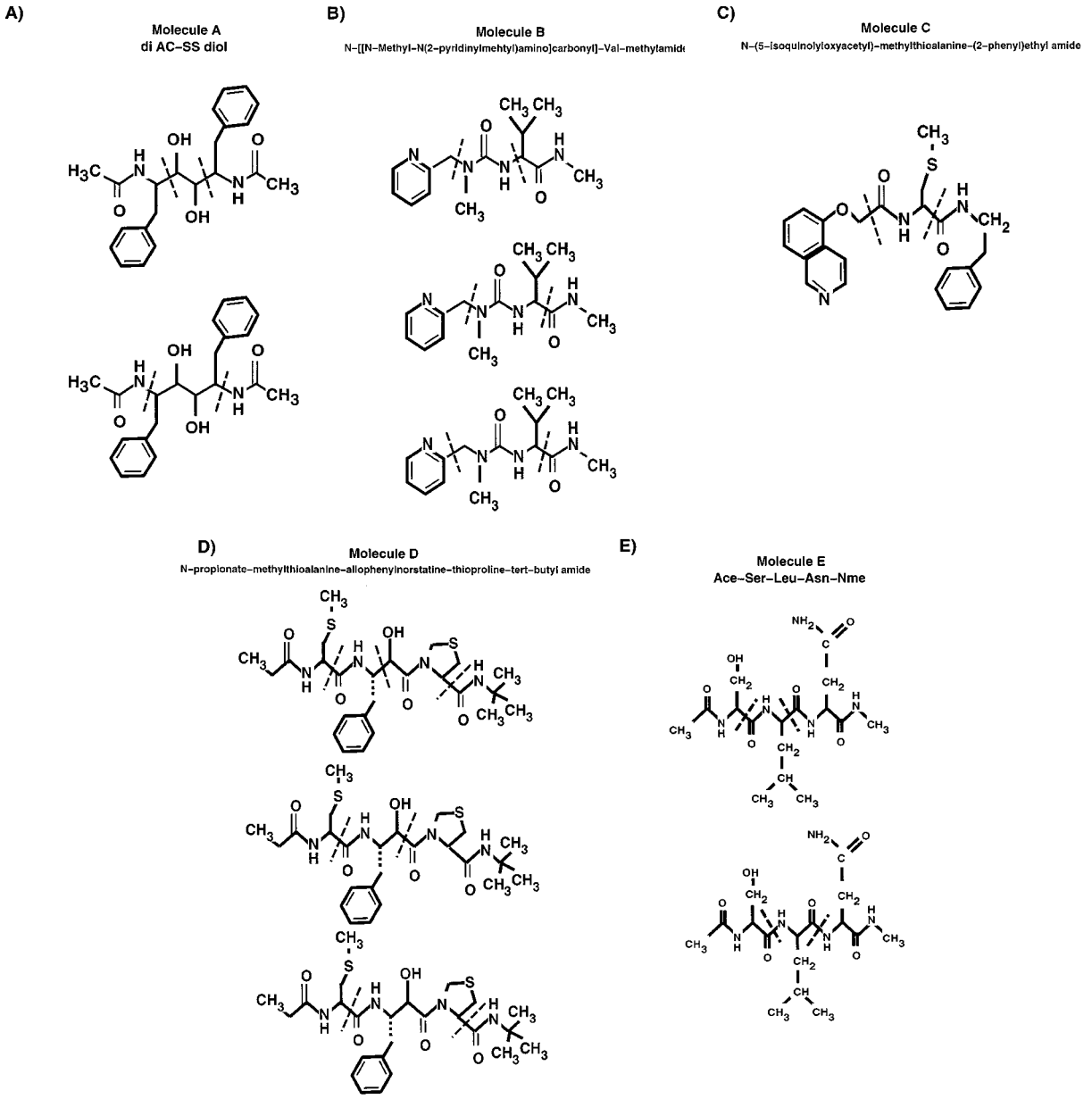

FIGURE 2. Five molecules (A–E), which are pieces of HIV-1 protease inhibitors, are used to study the fragment problem. The dashed lines represent the partitions used to generate the fragments. The combination of partitions for a molecule is referred to as the fragmentation scheme. In most cases, more than one scheme is studied.

the —H neutralization method is 1.4 kcal/mol, on average. Using the —CH_3 neutralization method,[21] it is 2.0 kcal/mol, on average, and the CEP fit method yields 1.3 kcal/mol, on average. The electrostatic potential was also calculated from the atomic partial charges of fragments with methyl terminal groups; however, the results of the fit of atomic partial charges to this combined electrostatic potential from —CH_3 terminal fragments were mixed and not as consistent as those using the —H terminal group.

ϕ^{whole} and $\phi^{reassoc}$ are the electrostatic potential due to partial charges from the whole molecule calculation and the fragment joined calculation, respectively. Differences in these values at a single point usually range from 0 to 6 kcal/mol in absolute value for the best schemes of each test molecule (italicized values in Table I), with the

TABLE I.
Comparison of Electrostatic Potentials of Reassociated Molecules with Those of Whole Molecules for Neutralization and CEP Fit Methods.

Fragmentation scheme[a]	Reassociation method	EP RMS diff (kcal / mol)
(A) Test molecule A (di AC-SS diol) (N = 2046[b])		
CT + CT, CT + CT	—H, neutralization	1.219
CT + CT, CT + CT	CEP fit	*0.914*
CT + CT, CT + CT	—CH_3, neutralization	1.853
CT + N, CT + N	—H, neutralization	4.771
CT + N, CT + N	CEP fit	2.691
CT + N, CT + N	—CH_3, neutralization	3.636
(B) Test molecule B (N = [[*N*-methyl-*N*-(2-pyridinylmethyl)amino]carbonyl]-Val-methylamide) (N = 1641)		
CT + N, CT + N	—H, neutralization	2.660
CT + N, CT + N	CEP fit	1.725
CT + N, CT + N	—CH_3, neutralization	3.577
CT + N, CT + C	—H, neutralization	1.607
CT + N, CT + C	CEP fit	*1.296*
CT + N, CT + C	—CH_3, neutralization	2.426
CT + CR, CT + C	—H, neutralization	1.776
CT + CR, CT + C	CEP fit	1.451
CT + CR, CT + C	—CH_3, neutralization	1.688
(C) Test molecule C (*N*-(5-isoquinolyloxyacetyl)-methylthioalanine-(2-phenyl)ethyl amide) (N = 1973)		
CT + C, CT + C	—H, neutralization	*1.435*
CT + C, CT + C	CEP fit	1.665
CT + C, CT + C	—CH_3, neutralization	2.385
(D) Test molecule D (*N*-propionate-methylthioalanine-allophenylnorstatine-thioproline-tert-butylamide) (N = 2413)		
CT + C, CT + CT, C + CP	—H, neutralization	2.961
CT + C, CT + CT, C + CP	CEP fit	1.870
CT + C, CT + C	—H, neutralization	1.839
CT + C, CT + C	CEP fit	*1.289*
CT + C, C + CP	—H, neutralization	1.793
CT + C, C + CP	CEP fit	1.495
(E) Test molecule E (Ace–Ser–Leu–Asn–Nme) (N = 1975)		
CT + C, CT + C	—H, neutralization	1.184
CT + C, CT + C	CEP fit	*1.155*
C + N, C + N	—H, neutralization	2.500
C + N, C + N	CEP fit	3.538

[a]Fragmentation schemes are shown in Figure 2.

[b]EP RMS diff = $\sqrt{\sum_{i=1}^{N}(\phi_i^{whole} - \phi_i^{reassoc})^2 / N}$, where, for N grid points, ϕ_i^{whole} is the electrostatic potential at the ith grid point due to the partial charges calculated for the unfragmented molecule, and $\phi_i^{reassoc}$ is the electrostatic potential at the ith grid point due to the partial charges calculated for the reassembled molecule.

largest deviations appearing near the region of fragmentation. Such deviations appear for only a few points out of a total of more than 1000.

Finally, we note that the partial charges of the whole molecule are obtained by fitting the EP calculated from the electron density. The quality of this fit is also given by an EP RMS difference. For the set of test molecules in Figure 2, these values range from 1.0 to 1.3 kcal/mol.

DIPOLE MOMENT AND HYDRATION ENERGY COMPARISONS

In addition to comparing the electrostatic potentials of the reassociated and whole molecules, dipole moments and hydration energies were also calculated. Table II shows dipole moments and free energies of hydration for the molecules and fragmentation schemes already introduced (Fig. 2)

TABLE II.
Comparison of Properties of Reassociated Molecules With Those of the Whole Molecule.

Fragmentation	Reassoc. method	$\mu^{reassoc}$ (Debye)	$\Delta\mu$[b] (Debye)	angle[c] (degrees)	$\frac{\vec{\mu}^{whole}\cdot\vec{\mu}^{reassoc}}{(\mu^{whole})^2}$	$\Delta G_{solv}^{reassoc}$ (kcal / mol)	$\Delta G_{solv}^{whole} - \Delta G_{solv}^{reassoc}$ (kcal / mol)
(A) Test molecule A (di AC-SS diol) $\Delta G_{solv}^{whole} = -25.29$ kcal / mol, $\mu^{whole} = 5.5$ Debye							
CT + CT, CT + CT	—H, neut.	6.1	−0.6	1.9	1.1	−26.05	0.76
CT + CT, CT + CT	CEP fit	5.9	−0.4	1.2	1.1	−25.54	*0.25*
CT + CT, CT + CT	—CH_3, neut.	6.3	−0.8	3.4	1.1	−27.01	1.72
CT + N, CT + N	—H, neut.	6.5	−1.0	4.0	1.2	−33.19	7.90
CT + N, CT + N	CEP fit	7.0	−1.5	4.6	1.3	−29.00	3.71
CT + N, CT + N	—CH_3, neut.	6.5	−1.0	4.9	1.2	−30.06	4.77
(B) Test molecule B (*N*-[[*N*-methyl-*N*-(2-pyridinylmethyl)amino]carbonyl]-Val-methylamide) $\Delta G_{solv}^{whole} = -12.23$ kcal / mol, $\mu^{whole} = 2.2$ Debye							
CT + N, CT + N	—H, neut.	2.3	−0.1	39.7	0.8	−13.55	1.32
CT + N, CT + N	CEP fit	2.0	0.2	23.6	0.8	−12.04	*−0.19*
CT + N, CT + N	—CH_3, neut.	2.2	0.0	43.5	0.7	−13.22	0.99
CT + N, CT + C	—H, neut.	1.9	0.3	11.4	0.9	−12.45	0.22
CT + N, CT + C	CEP fit	2.3	−0.1	2.5	1.1	−11.81	−0.42
CT + N, CT + C	—CH_3, neut.	2.1	0.1	28.7	0.8	−12.60	0.37
CT + CR, CT + C	—H, neut.	2.3	−0.1	10.0	1.0	−11.84	−0.39
CT + CR, CT + C	CEP fit	2.9	−0.7	5.5	1.3	−11.56	−0.67
CT + CR, CT + C	—CH_3, neut.	2.1	0.1	7.3	1.0	−11.77	−0.46
(C) Test molecule C (*N*-(5-isoquinolyloxyacetyl)-methylthioalanine-(2-phenyl)ethyl amide) $\Delta G_{solv}^{whole} = -16.74$ kcal / mol, $\mu^{whole} = 5.0$ Debye							
CT + C, CT + C	—H, neut.	5.5	−0.5	8.6	1.1	−17.03	*0.29*
CT + C, CT + C	CEP fit	5.3	−0.3	12.5	1.0	−17.03	*0.29*
CT + C, CT + C	—CH_3, neut.	5.6	−0.6	13.3	1.1	−18.63	1.89
(D) Test molecule D (*N*-propionate-methylthioalanine-allophenylnorstatine-thioproline-tert-butylamide) $\Delta G_{solv}^{whole} = -28.82$ kcal / mol, $\mu^{whole} = 3.1$ Debye							
CT + C, CT + CT, C + CP	—H, neut.	4.3	−1.2	19.8	1.3	−32.69	3.87
CT + C, CT + CT, C + CP	CEP fit	3.5	−0.4	4.3	1.1	−30.95	2.13
CT + C, CT + C	—H, neut.	3.7	−0.6	13.4	1.2	−29.73	0.91
CT + C, CT + C	CEP fit	2.9	0.2	10.7	0.9	−29.04	*0.22*
CT + C, C + CP	—H, neut.	3.7	−0.6	11.3	1.2	−31.46	2.64
CT + C, C + CP	CEP fit	3.3	−0.2	7.8	1.1	−31.16	2.34
(E) Test molecule E (Ace−Ser−Leu−Asn−Nme) $\Delta G_{solv}^{whole} = -35.11$ kcal / mol, $\mu^{whole} = 9.8$ Debye							
CT + C, CT + C	—H, neut.	10.0	−0.2	2.5	1.0	−35.52	0.41
CT + C, CT + C	CEP fit	10.3	−0.5	1.9	1.1	−35.45	*0.34*
C + N, C + N	—H, neut.	8.9	0.9	7.6	0.9	−32.42	−2.69
C + N, C + N	CEP fit	8.4	1.4	11.3	0.8	−30.57	−4.54

[a] Fragmentation schemes are shown in Figure 2.
[b] $\Delta\mu = \mu^{\text{whole}} - \mu^{reassoc}$.
[c] The angle is between $\vec{\mu}^{whole}$ and $\vec{\mu}^{reassoc}$.

using the CEP fit, —H neutralization, and —CH_3 neutralization methods of reassociation. Because properties such as free energies of hydration are sensitive not only to the magnitude of the dipole moment, but also to its direction,[4] a normalized vector inner product between the dipole moment vector of the whole molecule $\vec{\mu}^{whole}$ and that of the reassociated molecule $\vec{\mu}^{reassoc}$ is used to evaluate the reassociated molecule charge distribution: $\vec{\mu}^{whole} \cdot \vec{\mu}^{reassoc}/(\mu^{whole})^2$. In this way, if the two vectors are perfectly aligned and of the same magnitude, the value of the expression is one. Table II shows that the dipole moments of the reassociated molecules are comparable to those of the whole molecules, as $\vec{\mu}^{whole} \cdot \vec{\mu}^{reassoc}/(\mu^{whole})^2$ deviates from 1.0 by a maximum of ± 0.3. For a single molecule and fragmentation scheme, Table II shows that this expression is often the same or deviates very little for different methods. Thus the dipole moment vector is not a sensitive measure for distinguishing between various reassociation methods for a particular fragmentation scheme. On the other hand, the free energy of hydration comparisons are more sensitive and follow very closely the rankings found using the EP RMS comparisons.

Table II shows that the use of the CEP fit method for partitions between carbon atoms yields free energies of hydration for the reassociated molecule which are closest to those of the whole molecule. These numbers range in absolute value from 0.2 kcal/mol to 0.3 kcal/mol and are shown in italics in Table II. The smallest absolute value of the differences in hydration energies of the reassociated and whole molecules for each of the five test molecules using the —H neutralization method are 0.5 kcal/mol, on average. For the —CH_3 neutralization method, they are 1.3 kcal/mol, on average,[21] and for the CEP fit method, 0.3 kcal/mol, on average. The reassociated charge distributions reproduce the whole molecules hydration energies better than the whole molecule surface point EP values.

INTERACTION ENERGY COMPARISONS

Interaction energies are a direct function of the molecular charge distribution. This property is rarely readily available in the evaluation of atomic partial charges. Since the test molecules in this study are constitutive segments of inhibitors of HIV PR, we have a unique opportunity to check the effect of fragmentation schemes and reassociation methods on the interaction energy of the test molecule with HIV PR. The last column of Table III shows the differences between interaction energies calculated with the whole molecule charge distribution and those calculated with the reassociated molecule charge distribution. The smallest absolute value of these differences for each molecule is 0.8 kcal/mol, on average, for —H neutralization, 1.7 kcal/mol,[21] on average, for —CH_3 neutralization, and 0.7 kcal/mol, on average, for the CEP fit method. For three of the five molecules, —H neutralization gives the smallest differences, whereas, the CEP fit method is comparable, albeit slightly better, on average.

CEP FIT METHOD (VERSION 2)

Version 2 of the CEP fit method was tested for molecule A using scheme CT+ CT, CT+ CT and molecule B using scheme CT+ N, CT+ N. In this version, the fragment EP values are calculated on the whole molecule surface using DFT/DZVPD, and no fragment partial charges are required. The fragment EP values are simply summed and the whole molecule partial charges fit to this CEP. The EP RMS difference was improved by 0.15 kcal/mol for molecule A and 0.19 kcal/mol for molecule B. Because these changes are not significant, this version of the CEP fit method was not repeated for the other molecules and schemes. The improvement reflects the small error in the fragment partial charge distribution.

OVERLAP PARTITIONING

The overlap partitioning scheme was tested for molecules A and E. These molecules were chosen because they yield the best results for EP RMS comparisons. Two overlapping schemes were tested for molecule A as shown in Figure 3A and B. In one case, the overlap is the diol group only; in the other case, the overlap is extended and includes both the diol group and the Phe side chains. The EP RMS difference for the diol overlap is 1.5 kcal/mol, which is not an improvement over the other methods. For the diol and Phe side chains overlap, the EP RMS difference is 2.5 kcal/mol. The error introduced by cutting the CT—N bond is lessened in this case by the use of overlapping fragments, but the CEP fit method (Table IA), with fragmentation scheme CT+ CT, CT+ CT and an EP RMS of 0.9 kcal/mol, still gives better results than either of these overlapping fragment calculations. The calculated dipole moment ratio, $\vec{\mu}^{whole} \cdot$

TABLE III.
Comparison of HIV PR — Reassociated Molecule Interaction Energies With Those of the Whole Molecules for Neutralization and EP Fit Methods.

Fragmentation scheme[b]	Reassociation method	Interaction energy[c] (kcal / mol)	Difference[d] (kcal / mol)
(A) Test molecule A (di AC-SS diol), interaction energy[a] = −109.3 kcal / mol			
CT + CT, CT + CT	—H, neutralization	−111.2	1.9
CT + CT, CT + CT	CEP fit	−110.4	*1.1*
CT + CT, CT + CT	—CH_3, neutralization	−112.5	3.2
CT + N, CT + N	—H, neutralization	−113.3	4.0
CT + N, CT + N	CEP fit	−114.5	5.2
CT + N, CT + N	—CH_3, neutralization	−111.2	1.9
(B) Test molecule B (*N*-[[*N*-methyl-*N*-(2-pyridinylmethyl)amino]carbonyl]-Val-methylamide), interaction energy = −73.8 kcal / mol			
CT + N, CT + N	—H, neutralization	−73.9	*0.1*
CT + N, CT + N	CEP fit	−72.5	−1.3
CT + N, CT + N	—CH_3, neutralization	−73.3	−0.5
CT + N, CT + C	—H, neutralization	−73.3	−0.5
CT + N, CT + C	CEP fit	−72.5	−1.3
CT + N, CT + C	—CH_3, neutralization	−73.2	−0.6
CT + CR, CT + C	—H, neutralization	−73.7	*−0.1*
CT + CR, CT + C	CEP fit	−73.2	−0.6
CT + CR, CT + C	—CH_3, neutralization	−72.4	−1.4
(C) Test molecule C (*N*-(5-isoquinolyloxyacetyl)-methylthioalanine-(2-phenyl)ethyl amide), interaction energy = −75.9 kcal / mol			
CT + C, CT + C	—H, neutralization	−75.9	*0.0*
CT + C, CT + C	CEP fit	−75.7	−0.2
CT + C, CT + C	—CH_3, neutralization	−78.5	2.6
(D) Test molecule D (*N*-propionate-methylthioalanine-allophenylnorstatine-thioproline-tert-butylamide), interaction energy = −143.2 kcal / mol			
CT + C, CT + CT, C + CP	—H, neutralization	−145.9	2.7
CT + C, CT + CT, C + CP	CEP fit	−142.3	−0.9
CT + C, CT + C	—H, neutralization	−144.3	1.1
CT + C, CT + C	CEP fit	−141.6	−1.6
CT + C, C + CP	—H, neutralization	−144.2	1.0
CT + C, C + CP	CEP fit	−143.0	*−0.2*
(E) Test molecule E (Ace −Ser −Leu −Asn −Nme), interaction energy = −115.3 kcal / mol			
CT + C, CT + C	—H, neutralization	−116.4	*1.1*
CT + C, CT + C	CEP fit	−116.6	1.3
C + N, C + N	—H, neutralization	−113.8	−1.5
C + N, C + N	CEP fit	−108.3	−7.0

[a] This interaction energy is for the whole molecule−enzyme complex.
[b] Fragmentation schemes are shown in Figure 2.
[c] This interaction energy is for the reassociated molecule−enzyme complex.
[d] The difference is between the enzyme−whole molecule interaction energy and the enzyme-reassociated molecule interaction energy.

$\vec{\mu}^{reassoc}/(\mu^{whole})^2$, was 1.1 for the diol overlap and 1.0 for the extended overlap. The hydration free energies are larger than that calculated for the whole molecule by 2.1 kcal/mol for the diol overlap and 3.4 kcal/mol for the extended overlap as compared with 0.25 kcal/mol for the CEP fit method and 0.76 kcal/mol using —H neutralization (Table IA). The enzyme−molecule A interac-

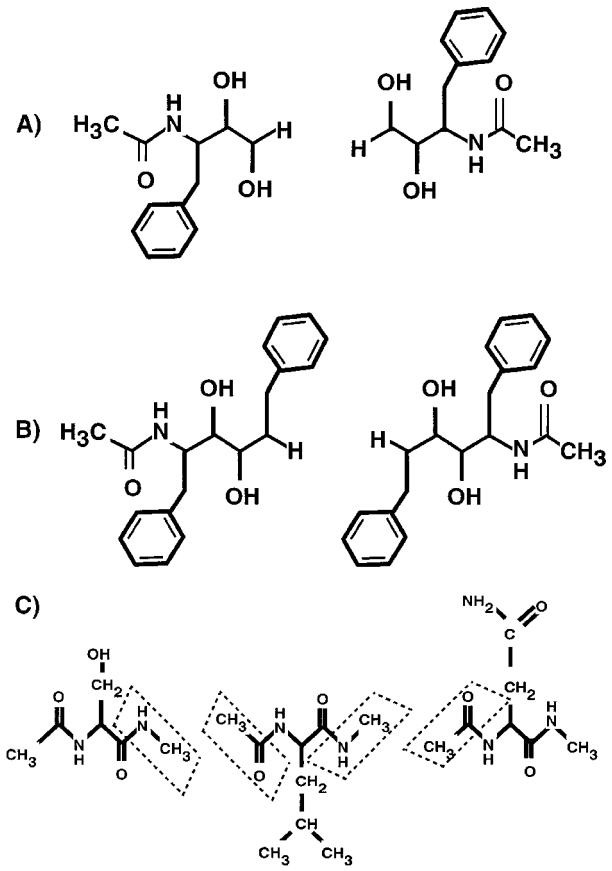

FIGURE 3. (A) The diol group is overlapping for these two fragments of molecule A. (B) The overlap is extended and includes the diol group and the Phe side chains. (C) These residues are used to test the average overlap method of reassociation and the neutral blocking group method for Ace–Ser–Leu–Asn–Nme (molecule E). Only in the neutral blocking group method are the groups denoted by dashed parallelograms constrained to be neutral.

tion energies are larger than that with the whole molecule by 3.0 kcal/mol and 3.2 kcal/mol, for the diol overlap and extended overlap, respectively. The corresponding value for the CEP fit method is 1.1 kcal/mol and for —H neutralization is 1.9 kcal/mol (Table III, part A). For molecule A, the average overlap method yields worse results than —H neutralization and CEP fit.

Another test of the average overlap method is performed on a peptide. The fragments chosen are each amino acid of molecule E, with the addition of acetyl and N-methyl blocking groups (Fig. 3C). An example of the overlap using molecule E is as follows. Upon reassociation, the methyl carbon of the acetyl blocking group of leucine overlaps the alpha carbon of the serine residue, and the alpha carbon of the leucine residue is overlapped by terminal methyl carbons in blocking groups of both the serine and asparagine residues. All of the intervening backbone atoms overlap. The EP RMS difference for this reassociation method and fragmentation scheme is 1.01 kcal/mol, smaller by less than 0.2 kcal/mol than the CEP fit method of Table I, part E. In this case, the overlapping fragment scheme and averaging method show only a slight advantage as evaluated by the EP RMS difference. The dipole moment ratio is 1.0 and the free energy of hydration of the reassociated molecule is larger than that of the whole molecule by 0.14 kcal/mol, as compared with 0.34 kcal/mol for the CEP fit method and 0.41 kcal/mol for the —H neutralization method. However, the enzyme–molecule E interaction energy is larger than the enzyme–original molecule interaction energy by 1.9 kcal/mol. The corresponding values for the —H neutralization method and for the CEP fit method are 1.1 kcal/mol and 1.3 kcal/mol, respectively. For molecule E, the average overlap method yields results comparable to —H neutralization and CEP fit; however, the interaction energy comparison is worse for the average overlap method. Based on these results for molecules A and E, the —H neutralization and the CEP fit methods are likely to be more reliable for reassociation than the average overlap method.

NEUTRAL BLOCKING GROUPS

The use of neutral blocking groups has been employed for amino acid charge calculations.[9] Thus, we examine this method on test molecule E which is a peptide. The neutral blocking groups are designated by the dashed parallelograms in Figure 3C. The EP RMS difference for this method is 1.23 kcal/mol; the dipole moment ratio is 1.0; the free energy of solvation is −35.54 kcal/mol; and the interaction energy is −116.8 kcal/mol. All of these values are comparable to those (parts E of Tables I, II, and III) of the CEP fit and —H neutralization methods.

Application to HIV-1 Protease Diastereomer Inhibitors

Having studied five test molecules that were obtained by truncation of HIV PR, we now extend our calculations to the full size inhibitors (Fig. 4). Interaction energies of three diastereomers are cal-

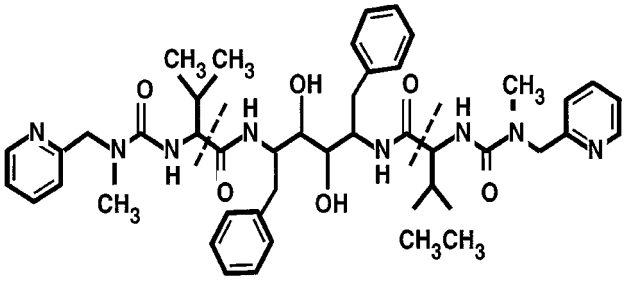

FIGURE 4. The RR, RS, and SS inhibitors differ only in the chirality at the diol (the diol comprises the two central hydroxyl groups). The partitions used in the partial charge calculation are represented by dashed lines.

culated and compared to binding affinity experiments. These three diastereomers are RR (A76889), RS (A77003), and SS (A76928) with inhibition constants[17] K_i of 112 pM, 12 pM, and 11 pM, respectively, and corresponding RTln(K_i) values of -14.1, -15.5, and -15.6 kcal/mol at 37°C.[17] We chose to study diastereomers because the effects of entropy and hydration on binding are expected to differ very little for the three. A complete calculation of the free energy of binding is beyond the scope of this article, but the calculated interaction energies should reproduce the trend: RR > RS $\simeq$ SS.

Partitions for the partial charge calculation are placed between each valine C_α and carbonyl carbon as shown by the dashed lines in Figure 4. Interaction energies (Table IV) were calculated using CHARMm 22 and AMBER 4.0 (including crystallographic water molecules) with DFT/DZVPD CEP fit reassociated partial charges for the inhibitors. The order of binding is reproduced by both CHARMm and AMBER. In CHARMm, the van der Waals energies are the dominant term. In AMBER, the coulomb energies dominate slightly.

TABLE IV.
HIV-1 Protease-Inhibitor Interaction Energies.

Complex	vdW[a]	Coulomb	Total
CHARMm 22			
RR	−104.12	−59.01	−163.13
RS	−104.51	−85.25	−189.76
SS	−113.73	−76.96	−190.69
AMBER 4.0			
RR	−67.34	−68.29	−135.63
RS	−71.58	−96.40	−167.98
SS	−73.47	−93.03	−166.50

[a] van der Waals energy.

This agreement of the interaction energies with the trend in binding may be entirely fortuitous. The effect of hydration should reduce the magnitude of these numbers considerably, and efforts to adapt the use of the boundary element method to calculate hydration energies for such large systems are underway.

Conclusions

Earlier work[3,4] shows that the use of an extended basis set with diffuse functions is required to find partial charges which reproduce experimental dipole moments well. The calculation of partial charges for molecules on the order of 50–60 nonhydrogen atoms using this basis set is at present intractable. This study represents our first step in understanding the errors introduced by partitioning and reassociation of a molecule in calculating atomic partial charges using a large basis set. We have compared various fragmentation schemes and reassociation methods to obtain atomic partial charges for these molecules. This work has shown that errors in the electrostatic potential introduced by fragmenting a tripeptide at the peptide bond are two to five times larger than for the alpha carbon—carbonyl carbon bond. This was anticipated due to the polar nature of the peptide bond. In general, the expectation, that errors introduced by partitioning of carbon—carbon bonds would be smaller than those from partitioning of carbon—nitrogen bonds, was realized. Overall, the CEP fit method works slightly better than the —H neutralization method. For peptides, the neutral blocking group method[9] was also tested and found to be comparable to the CEP fit method.

The results indicate that our solvation calculations are less sensitive to errors in the charge distribution than calculations of the electrostatic potential at the van der Waals surface shells. The sensitivity of interaction energy calculations to these errors is more selective. That is, not only do the portions of the charge distribution at the ligand–target interface dominate the result, but also, certain portions of this interface, such as regions containing highly charged groups, can magnify errors in the charge distribution. As the largest errors are usually near the partitions, the partitions can sometimes be chosen based on knowledge of the target structure. For example, with the target HIV PR, the ligand partitions should be made away from the active-site aspartic acid residues in the bound structure. Further reduction of these

errors will come from a better representation of interfragment polarization effects and will most likely be incorporated using some type of CEP fit method.

Acknowledgments

We thank Dr. John Erickson and Dr. T. N. Bhat for crystallographic coordinates. We thank Mr. Robert Lebherz for a prompt upgrade of Gaussian 92 to include the DFT code. We thank Dr. Jack Collins for sharing his expertise with AMBER, Dr. Greg Tawa for insightful discussions, and Ms. Lucyna Lubkowska for chemical nomenclature. Special thanks are extended to the staff of the Biomedical Supercomputing Center, FCRDC, Frederick, MD, for their assistance and for access to the Cray Y-MP supercomputer. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

References

1. B. H. Besler, K. M. Merz, Jr., and P. A. Kollman, *J. Comput. Chem.*, **11**, 431 (1990).
2. A. A. Rashin, M. A. Bukatin, J. Andzelm, and A. T. Hagler, *Biophys. Chem.*, **51**, 375 (1994).
3. A. A. Rashin, L. Young, I. A. Topol, and S. K. Burt, *Chem. Phys. Lett.*, **230**, 182 (1994).
4. A. A. Rashin, L. Young, and I. A. Topol, *Biophys. Chem.*, **51**, 359 (1994).
5. M. N. Bellido and J. A. C. Rullmann, *J. Comput. Chem.*, **10**, 479 (1989).
6. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, **5**, 129 (1984).
7. K. M. Merz, Jr., *J. Comput. Chem.*, **13**, 749 (1992).
8. Averaging the partial charges of the overlapping atoms yields an excess charge on the molecule on the order of hundredths of atomic units. Thus, the partial charges of all the atoms in the molecule must be altered by the amount of excess charge divided by the number of atoms—a value on the order of 10^{-4} atomic units.
9. P. Cieplak, W. D. Cornell, C. Bayly, and P. A. Kollman, *J. Comput. Chem.*, **16**, 1357 (1995).
10. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. W. Wong, J. B. Foresman, M. A. Robb, M. Head-Gordon, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian 92/DFT, revision G.2, Gaussian, Inc., Pittsburgh, PA, 1993.
11. A. D. Becke, *J. Chem. Phys.*, **93**, 5648 (1993).
12. J. P. Perdew, *Phys. Rev. B*, **33**, 8822 (1986), erratum in J. P. Perdew, *Phys. Rev. B*, **38**, 7406 (1986).
13. The diffuse *d*-exponent for the sulphur atom was obtained using simple sulfur-containing molecules. Only the diffuse *d*-exponent in the DZVPD basis set of sulfur was varied to obtain the minimum energy for H_2S. Dipole moments for H_2S, CH_4S, and $(CH_3)_2SO$ calculated with this optimal DZVPD basis set were within 0.1 Debye from experimental vacuum dipole moments.
14. J. Brady, B. R. Brooks, R. E. Bruccoleri, A. Brunger, M. J. Field, T. Ichiye, L. Nilsson, B. D. Olafson, D. Perahya, C. B. Post, W. E. Reiher II, D. J. States, S. Swaminathan, B. Tidor, and M. Karplus, CHARMm, version 22, Copyright 1984, 1987, 1991, President and Fellows of Harvard College.
15. B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).
16. D. A. Pearlman, D. A. Case, J. C. Caldwell, G. L. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman, AMBER 4.0, University of California, San Francisco, 1991.
17. M. V. Hosur, T. N. Bhat, D. J. Kempf, E. T. Baldwin, B. Liu, S. Gulnik, N. E. Wideburg, D. W. Norbeck, K. Appelt, and J. W. Erickson, *J. Am. Chem. Soc.*, **116**, 847 (1994).
18. E. T. Baldwin, T. N. Bhat, S. Gulnik, B. Liu, I. A. Topol, Y. Kiso, T. Mimoto, H. Mitsuya, and J. W. Erickson, *Structure*, **3**, 581 (1995).
19. A. L. Swain, M. M. Miller, J. Green, D. H. Rich, J. Schneider, S. B. H. Kent, and A. Wlodawer, *Proc. Natl. Acad. Sci. USA*, **87**, 8805 (1990).
20. Per unit charge is assumed.
21. Based on the evaluations for molecules A, B, and C, —CH_3 neutralization was not tested on molecules D and E. Thus, the average is for three molecules.